Data on Quantitative Microbial Elemental Cycling (QMEC) primer design and validation

Zheng, Bang-Xiao

2019-04

Zheng , B-X & Ding , K 2019 , ' Data on Quantitative Microbial Elemental Cycling (QMEC) primer design and validation ' , Data in Brief , vol. 23 , 103820 . https://doi.org/10.1016/j.dib.2019.103820

http://hdl.handle.net/10138/304413
https://doi.org/10.1016/j.dib.2019.103820

Data in brief 23 (2019) 103820

Contents lists available at ScienceDirect

Data in brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Data on Quantitative Microbial Elemental Cycling (QMEC) primer design and validation

Bang-Xiao Zheng [a, b], Kai Ding [a, *]

[a] Key Laboratory of Urban Environment and Health, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021, PR China
[b] Faculty of Biological and Environmental Sciences, Ecosystems and Environment Research Programme, University of Helsinki, Niemenkatu 73, Lahti 15140, Finland

ARTICLE INFO

Article history:
Received 26 December 2018
Received in revised form 20 February 2019
Accepted 4 March 2019
Available online 12 March 2019

ABSTRACT

This work included the primer design details of Quantitative Microbial Elemental Cycling (QMEC) method, a high-throughput qPCR method for microbial carbon (C), nitrogen (N), phosphorus (P), sulfur (S) and methane metabolism potential detection and assessment. We designed 36 novel primers based on their amino acid sequences. Via illumina sequencing technology, their phylogenetic taxonomy was identified and analyzed to validate the primer specificity.

Specifications table

Subject area	*Biology*
More specific subject area	*Molecular Ecology*
Type of data	*Tables and figures*
How data was acquired	*Illumina sequencing, high-throughput qPCR*
Data format	*Raw reads and analysis*
Experimental factors	*Soil and sediment samples*
Experimental features	*New primer design based on protein sequences and high-through sequencing*
Data source location	*Hailun, Heilongjiang, China (47°26′N, 126°38′E); Qiangtang River, Hangzhou, China (30°39′N, 120°52′E).*

* Corresponding author.
E-mail address: kding@iue.ac.cn (K. Ding).

https://doi.org/10.1016/j.dib.2019.103820

Data accessibility	*All obtained sequences were submitted to the National Center for Biotechnology Information Sequence Read Archive (SRA) with the accession numbers SRP107153 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA386825/) and SRP107154 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA386826/).*
Related research article	*1. Bangxiao Zheng, Yongguan Zhu, Jordi Sarrdans, Josep Peñuelas, Jianqiang Su. (2018) QMEC: a tool for highthroughput quantitative assesement of microbial functional potential in C, N, P, and S biogeochemical cycling. Sci China Life Sci 61, http://doi.org/10.1007/s11427-018-9364-7.* *2. Yong-Guan Zhu, Bang-Xiao Zheng, Jian-Qiang Su, Hu Li, Huai-Ying Yao. High-throughput detection primer and detection method for microbial carbon, nitrogen, phosphorus and sulfur functional genes. Chinese Patent: CN201610537546.*

Value of the Data

- Multiple useful primers (designed or cited) were merged into one detection method will provide a convenient way for researchers to conduct a comprehensive investigation of microbial functional gene abundance regarding C, N, P and S cycling.
- The data describes the details of primer designs, which would be supportive for other researchers to make repetitive verification.
- The data retrieved from high-throughput qPCR could be used in comparison with other related methodologies.
- The data from soil and sediment could be used to make comparison with reported studies.

1. Data

The data included 36 primer design and validation by illumina sequencing and qPCR detection. Fig. 1 gave a heatmapping species-summary based on sequencing results of 36 designed primers which was aligned with RefSeq database. Two samples, soil and sediment, were given in Fig. 1a and b,

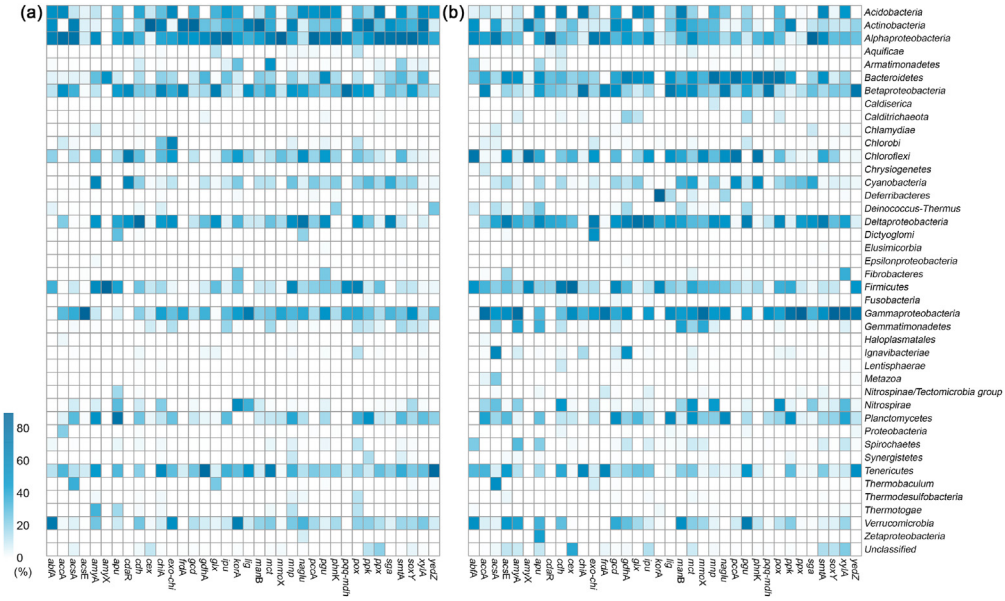

Fig. 1. The species coverage percentage in phylum or class level based on the sequencing results of 36 designed primers from soil (a) and sediment samples (b).

Table 1
Diversity indices for the functional genes using the designed primer pairs summarized at phylum level.

	Soil					Sediment				
	Shannon	Simpson	Inverse Simpson	Species number	Evenness	Shannon	Simpson	Inverse Simpson	Species number	Evenness
*abf*A	1.598	0.746	3.944	14	0.606	1.797	0.774	4.419	13	0.700
*acc*A	1.368	0.619	2.628	23	0.436	1.697	0.699	3.327	30	0.499
*acs*A	1.519	0.671	3.036	23	0.484	2.127	0.820	5.566	26	0.653
*amy*A	0.140	0.040	1.042	17	0.050	2.083	0.836	6.091	20	0.695
*acs*E	2.237	0.867	7.536	24	0.704	1.706	0.653	2.882	23	0.544
*amy*X	0.363	0.208	1.263	2	0.524	1.181	0.636	2.745	8	0.568
Apu	1.604	0.614	2.588	17	0.566	2.455	0.886	8.755	22	0.794
*cda*R	1.858	0.756	4.094	22	0.601	0.483	0.199	1.249	6	0.270
Cdh	1.339	0.567	2.309	17	0.473	1.858	0.756	4.094	22	0.601
Cex	0.355	0.116	1.131	14	0.134	1.339	0.567	2.309	17	0.473
*chi*A	2.033	0.816	5.436	26	0.624	1.389	0.689	3.211	16	0.501
exo-chi	2.036	0.835	6.052	15	0.752	1.533	0.699	3.325	16	0.553
*frd*A	1.137	0.606	2.538	6	0.635	1.410	0.728	3.674	15	0.521
Gcd	0.889	0.354	1.548	18	0.308	2.479	0.901	10.052	28	0.744
*gdh*A	1.244	0.542	2.186	17	0.439	2.209	0.853	6.817	24	0.695
Glx	1.302	0.628	2.689	16	0.470	1.461	0.664	2.977	13	0.570
Ipu	1.223	0.480	1.922	15	0.452	1.842	0.781	4.563	20	0.615
*kor*A	2.240	0.846	6.512	19	0.761	0.689	0.362	1.569	11	0.287
Lig	1.454	0.695	3.275	9	0.662	1.859	0.788	4.715	16	0.670
*man*B	0.959	0.366	1.578	19	0.326	2.404	0.879	8.267	19	0.816
Mct	1.842	0.762	4.195	19	0.626	2.354	0.872	7.829	25	0.731
*mmo*X	0.586	0.237	1.311	18	0.203	1.675	0.675	3.079	19	0.569
Mnp	2.090	0.814	5.372	23	0.666	1.832	0.728	3.683	31	0.533
Naglu	1.431	0.590	2.437	22	0.463	1.792	0.780	4.546	13	0.699
pccA	1.250	0.490	1.959	18	0.433	1.302	0.663	2.964	14	0.494
Pgu	2.209	0.863	7.297	23	0.705	1.964	0.801	5.030	21	0.645
*phn*K	1.249	0.494	1.977	18	0.432	1.339	0.670	3.031	14	0.507
pqq-mdh	1.151	0.611	2.568	12	0.463	1.264	0.657	2.917	17	0.446
Pox	2.126	0.822	5.608	20	0.710	1.770	0.720	3.568	17	0.625
Ppk	1.765	0.708	3.430	21	0.580	1.649	0.717	3.529	16	0.595
Ppx	1.429	0.636	2.747	19	0.485	0.562	0.234	1.306	14	0.213
Sga	1.546	0.695	3.281	16	0.557	1.140	0.474	1.899	19	0.387
*smt*A	1.383	0.520	2.083	21	0.454	1.836	0.779	4.534	21	0.603
*sox*Y	1.458	0.581	2.389	20	0.487	0.803	0.287	1.403	19	0.273
*xyl*A	1.629	0.713	3.487	21	0.535	1.736	0.661	2.948	30	0.510
*yed*Z	0.972	0.402	1.671	15	0.359	1.492	0.717	3.532	14	0.566

respectively. Table 1 shows the biodiversity analysis based on taxonomic assignments of designed primers from soil or sediment samples. The raw reads of illumina sequencing based on soil and sediment samples could be respectively obtained from NCBI with accession numbers SRP107153 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA386825/)and SRP107154 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA386826/).

2. Experimental design, materials, and methods

In QMEC method, we have designed 36 primer pairs potentially amplifying C, N, P and S cycling genes [1]. Some of which are modified from previous reports, based on the conserved regions of amino acid sequences alignments. Among them, 9 primers were specific (*acs*A, *kor*A, *lig*, *mmo*X, *phn*K, *pqq-mdh*, *ppx*, *sox*Y and *yed*Z) and other 27 primers were degenerated (*abf*A, *acc*A, *acs*E, *amy*A, *amy*X, *apu*, *cda*R, *cdh*, *cex*, *chi*A, *exo-chi*, *frd*A, *gcd*,*gdh*A, *glx*, *ipu*, *man*B, *mct*, *mnp*, *naglu*, *pccA*, *pgu*, *pox*, *ppk*, *sga*, *smt*A and *xyl*A). All the primers were designed based on their amino acid sequence alignment by ClustalW2 [2]. The potential dimers and hairpins were screened and excluded by Primer Primier version 5.0 [3]. Each primer was limited within 20–24 bp length and 50–65 GC%. The QMEC will conduct a unified

amplification procedure, the annealing temperature of each gene should be close. The gene product was restricted between 200 and 500 bp length. The corresponding region of each designed primer was labelled in aligned protein sequences and it could be found in Supplementary Material 1 (Figures SM1-36).

We used samples originated from soil and sediment to validate the specificity of primers. The sample information and DNA sequencing method could be found in our previous paper [1]. Briefly, the samples were DNA-extracted and PCR-amplified. All 36 pairs successfully generated amplicons with the expected sizes from the genomic DNA of the environmental samples. These amplicons were then purified and high-throughput sequenced. The returned raw reads were filtered, processed and analyzed using QIIME program [4]. Operational taxonomic units (OTUs) were clustered by UCLUST at a 3% dissimilarity cutoff [5]. The representative sequences were retrieved and classified using RDP classifier [6].The filtered sequences were searched against the RefSeq database (ftp://ftp.ncbi.nlm.nih.gov/refseq/release) using Local BLAST 2.2.27 + (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/2.2.27/) with 10^{-5} e value. The result with highest score was accepted only. A summary of the alignment results, including matched count and percentage, are listed in Supplementary Material 2. The taxonomic affiliation of these genes differed significantly between the soil and sediment samples at the phylum level. The species coverage of 36 primer pairs is visualized in Fig. 1, which shows *Acidobacteria*, *Actinobactera*, *Alphaproteobacteria* and *Gammaproteobacteria* were dominate speices in both soil and sediment samples. Taxonomic diversity based on the Shannon, Simposon's index, which were commonly used in mathematical measure of species diversity in a community [7], were listed in Table 1.

Acknowledgments

This work was financially supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB15020402, XDB15020302) and the Natural Science Foundation of China (41571130063, 41430858).

Transparency document

Transparency document associated with this article can be found in the online version at https://doi.org/10.1016/j.dib.2019.103820.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.dib.2019.103820.

References

[1] Bangxiao Zheng, Yongguan Zhu, Jordi Sarrdans, Josep Peñuelas, Jianqiang Su, QMEC: a tool for highthroughput quantitative assesement of microbial functional potential in C, N, P, and S biogeochemical cycling, Sci. China Life Sci. 61 (2018).
[2] M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWillian, et al., Clustal W and clustal X version 2.0, Bioinfromatics 23 (2007) 2947–2948.
[3] S. Lalitha, Primer primier 5. biotech software & internet report: the computer software, J Sci 1 (2000) 270–272.
[4] J.G. Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F.D. Bushman, E.K. Costello, et al., QIIME allows analysis of high-throughput community sequencing data, Nat. Methods 7 (2010) 335–336.
[5] R.C. Edgar, Search and clustering orders of magnitude faster than BLAST, Bioinofrmatics 26 (2010) 2460–2461.
[6] M.N. Price, P.S. Dehal, A.P. Arkin, FastTree: computing large minimum evolution trees with profiles instead of a distance matrix, Mol. Biol. Evol. 26 (2009) 1641–1650.
[7] Keylock C.J. Simpson, Diversity and the Shannon-Wiener index as special cases of a generalized entropy, Oikos 109 (2005) 203–207.